Frank David Faulkner

OPTIMUM SHIP ROUTING.

UNITED STATES
NAVAL POSTGRADUATE SCHOOL
MONTEREY, CALIFORNIA

OPTIMUM SHIP ROUTING

BY
FRANK D. FAULKNER

RESEARCH PAPER No. 32

MAY 1962

Optimum Ship Routing

by

Frank D. Faulkner
Professor of Mathematics and Mechanics

U. S. Naval Postgraduate School
Research Paper No. 32

Sponsored by
Stanford Research Institute
Menlo Park, California
for
Fleet Numerical Weather Facility
Monterey, California

September 1962

FOREWORD

The principal purpose of this paper is to present methods for determining minimum-time ship routes on a digital computer. The paper was written in two parts. The first treats in detail a numerical method for determining the course which requires minimum time to go from one specified place to another when the speed of the ship is a known function of position, heading, and time.

The second part of the paper treats various related topics. The first is the corresponding problem in the simpler case wherein the speed is a function of position and heading but not of time. The second is a discussion of methods for obtaining the curves, called equal-time curves or isochrones, which yield the maximum distance the ship can attain at any particular time by choosing various courses. These are discussed for both the stationary and time-varying speed fields; they are rather important because they are easily understood and interpreted, and closely parallel numerical methods now in use. Third, the problem of effecting rendezvous between two ships is treated. Finally a brief discussion is given of various problems and difficulaties which may be encountered in computation.

It has been the intention particularly in the first part to give the numerical routines and associated discussions in sufficient detail that a person familiar with numerical methods and computers could immediately program and run the problem. The method is based on procedures which G. A. Bliss introduced in ballistics for calculating differentials. These are applied in a Newton-Raphson iteration to determine the course. It is the author's opinion that the major outstanding part of the ship-routing problem is the collection of reliable empirical data describing the speed of the ship.

It is hoped that this paper will serve as an introduction to the adjoint system, calculus of variations, and optimum control theory along lines which are currently being actively pursued in this country and in Russia, including a method of solution. The first part of this paper has been accepted for publication in NAVIGATION, Journal of the Institute of Navigation.

Contents

A GENERAL NUMERICAL METHOD FOR DETERMINING OPTIMUM SHIP ROUTES

Frank D. Faulkner*

A method is given for determining optimum ship routes on a digital computer. This paper is limited to the problem of determining minimum time courses when the speed is a known function of the position, heading, and time.

The method is based on Bliss's methods for calculating differentials, usually applied in a Newton iteration to determine the course. It is very general, but it requires a knowledge of Bliss's adjoint methods. The paper is selfcontained; a simple case is given in detail and others outlined. The advent of high-speed digital computers in the past few years now makes the solution of such problems feasible.

1. Statement of problem.

We shall take the equations of motion to be expressed in the form

$$(1)\qquad \begin{cases} \mathring{x} = v \cos p \\ \mathring{y} = v \sin p \end{cases}$$

where x, y are position coordinates, p is a control variable, $v = v(x,y,p,t)$ is a known function of period 2π in p with continuous derivatives, t is time, and the dot (°) over a variable indicates its time derivative. If x, y are rectangular coordinates, then v is the speed and p is the angle between the velocity vector and the x axis, the heading angle;

*Professor of Mathematics and Mechanics, U. S. Naval Postgraduate School

if they are, say, latitude and longitude, then there is no such simple interpretation.

We will consider first the problem of going from one specified point, (0,0), to another, (X,Y), with minimum time T. The problem is to choose the control variable p as a function of time to effect this, and to determine the corresponding curve in x,y,t space. In a later section general conditions will be given for minimum time courses, as in rendezvous, etc., and corner conditions which would be of interest for a sailboat which must tack.

2. Variational equations, Euler equations.

In this section some formulas for differentials are derived, by procedures which G. A. Bliss introduced in Ballistics. Let us consider two neighboring paths, whereon the values of p differ by an amount δp; it is assumed that $\int_0^T |\delta p|\, dt$ is small. Then the change $\delta x, \delta y$ in x,y satisfy the variational equations

$$(2)\quad \begin{cases} \delta\dot{x} = v_x \delta x \cos p + v_y \delta y \cos p + (v_p \cos p - v \sin p)\delta p \\ \delta\dot{y} = v_x \delta x \sin p + v_y \delta y \sin p + (v_p \sin p + v \cos p)\delta p; \end{cases}$$

subscripts here indicate partial derivatives, $v_x = \partial v/\partial x$, etc.

Let us multiply these two equations through by two new variables λ, μ (unspecified at present but to be identified at some stage as Lagrange multipliers), collect terms, and integrate to get

$$(3)\quad \int_0^T [\lambda(\delta\dot{x} - v_x \delta x \cos p - v_y \delta y \cos p - v_p \delta p \cos p + v \delta p \sin p)$$
$$+ \mu(\delta\dot{y} - v_x \delta x \sin p - v_y \delta y \sin p - v_p \delta p \sin p - v \delta p \cos p)]dt = 0.$$

We may integrate this by parts and rewrite it as

(4) $$[\lambda\delta x + \mu\delta y]_0^T = \int_0^T [\delta x(\dot{\lambda} + \lambda v_x \cos p + \mu v_x \sin p) + $$
$$+ \quad \delta y(\dot{\mu} + \lambda v_y \cos p + \mu v_y \sin p)$$
$$+ \delta p(\lambda\{v_p \cos p - v \sin p\} + \mu\{v_p \sin p + v \cos p\})]dt$$
$$- [v(\lambda \cos p + \mu \sin p)]_{t_1-}^{t_1+} \delta t_1 \quad ,$$

where t_1 is a symbol for any point or points where p is discontinuous as a function of time, a <u>steering corner</u>. Probably there will be no such point for ships and we shall drop the last term for the present. To simplify this equation, let us choose λ, μ as solutions to the system

(5) $$\begin{cases} \dot{\lambda} + \lambda v_x \cos p + \mu v_x \sin p = 0 \\ \dot{\mu} + \lambda v_y \cos p + \mu v_y \sin p = 0 \ , \end{cases}$$

so that the coefficients of $\delta x, \delta y$ in (4) vanish. This defines the system (5) which is <u>adjoint</u> to the system (3).

In the case of most interest, the initial values of x,y are assumed known; then $\delta x(0) = 0 = \delta y(0)$, and (4) becomes

(6) $$\lambda(T)\delta x(T) + \mu(T)\delta y(T) =$$
$$= \int_0^T [\lambda(v_p \cos p - v \sin p) + \mu(v_p \sin p + v \cos p)]\delta p \, dt$$
$$= \int_0^T \vec{\Lambda} \cdot \vec{v}_p \, \delta p \, dt,$$

where $\vec{\Lambda} = \lambda\vec{i} + \mu\vec{j}$ and $\vec{v} = v(\vec{i} \cos p + \vec{j} \sin p)$. This is the fundamental differential formula connecting the change in end values of x,y with the variation of the control variable.

It is a very general condition for any optimum that there is some solution $\vec{\Lambda}$ to the adjoint system, to be found in the solution of the problem, and for that solution, $\vec{v}$ must satisfy the condition

(7) $\vec{\Lambda}\cdot\vec{v} \;=\;$ extreme,

as a function of the control variable p at every point of the course. This condition implies that the coefficient of δp in (6) vanish

(8) $\vec{\Lambda}\cdot\vec{v}_p \;=\; 0,$

for that solution. Comment. The proof of this condition is essentially the proof of the fundamental lemma in the calculus of variations (see Courant*[1] p 200). Cases wherein it does not determine a unique path are easily constructed, but seem to be of little practical significance and will not be considered further.

Equations (5), (8) are called the <u>Euler equations</u> in calculus of variations. The curves whereon equations (1),(5),(8) are satisfied are called <u>extremals</u> of the family (1). Equation (7) or (8) may be replaced by the condition that p be chosen so that

(9) $\displaystyle\int_0^T \vec{\Lambda}\cdot\vec{v}\,dt = \text{extreme};$

this approach has received considerable attention recently through the work of the Russian mathematician Pontriagin [2] on control.

The various philosophies of approach represented by (7), (8),(9) are equivalent for this problem, and none eliminates the basic problem of solution; namely, conditions are given at two values of t, $(0,T)$, with the second one unknown, and a constant is needed to integrate equations (1),(5),(8). In the next sections a method of solution is given.

3. Differential formulas.

Let us consider a fundamental set of solutions for (5), say

---

*Numbers in brackets refer to references listed at the end of the paper.

$\lambda_1, \mu_1, \lambda_2, \mu_2$ taken so that at $t = 0$ these have the values 1,0,0,1, respectively. Then every set of solutions is a linear combination of these two. Now the fundamental equation (6) is linear and homogeneous in the pair λ, μ so that a multiplicative factor will drop out from the formulas for $\delta x, \delta y$. Hence we may express all solutions of interest in terms of a single parameter a

$$\text{(10)} \qquad \begin{cases} \lambda = \lambda_1 \cos a + \lambda_2 \sin a \\ \mu = \mu_2 \cos a + \mu_2 \sin a \end{cases}$$

in so far as the use of this formula is concerned. An extremal which starts at the origin at $t = 0$ is characterized by the value of a except for the terminal time T, and the problem of determining an optimum course reduces to that of finding a, T.

It is shown in the appendix that for extremals a differential change in a leads to a differential change $\delta x(T), \delta y(T)$ in x, y at T

$$\text{(11)} \begin{cases} \delta x = -J\mu(T)\delta a \\ \delta y = J\lambda(T)\delta a \end{cases}$$

where

$$\text{(12)} \qquad J = \frac{1}{\begin{vmatrix} \lambda_1 & \lambda_2 \\ \mu_1 & \mu_2 \end{vmatrix}_{t=T}} \int_0^T \frac{(v^2+v_p^{\,2})^{1/2}}{(\lambda^2+\mu^2)^{3/2}} \begin{vmatrix} \lambda_1 & \lambda_2 \\ \mu_1 & \mu_2 \end{vmatrix}^2 dt$$

If both a, T are changed the resulting differential changes in the terminal values are

$$\text{(13)} \qquad \begin{cases} \Delta x(T) = \mathring{x}(T)\delta T - J\mu(T)\delta a \\ \Delta y(T) = \mathring{y}(T)\delta T + J\lambda(T)\delta a \end{cases}$$

4. Computational routine.

Let us guess initially values for a, T, say a_0, T_0. Then integrate simultaneously the original equations, a fundamental set $\lambda_1, \mu_1, \lambda_2, \mu_2$ of solutions to the adjoint system (6), with λ, μ determined then by (10) and p determined by (8); calculate also the integral J.

A curve thus determined will not go to the desired point X, Y generally, but to, say, X_i, Y_i, where i is the iteration index. In equation (13), set

$$(14) \quad \begin{cases} \Delta x_i = X - X_i \\ \Delta y_i = Y - Y_i \end{cases}$$

and determine $\delta a_i, \delta T_i$, $a_{i+1} = a_i + \delta a_i$, $T_{i+1} = T_i + \delta T_i$.

Continue this until some convergence criterion is satisfied, say,

$$(X - X_i)^2 + (Y - Y_i)^2 < \varepsilon ,$$

where ε is a prescribed number.
Of course, if a, T are guessed too far from the correct values it may not converge; this has not been a problem so far.

5. General end conditions.

In the most general case we may have a function $g(x,y,t)$ to be minimized, or maximized at the terminal point, subject to N constraints of the form

$$h_n(x,y,t) = 0, \qquad n = 1,\ldots N.$$

There may be none, one, or two of these ($N = 0, 1,$ or 2). In this case the conditions on the end values, called the transversal conditions, which must be satisfied for a stationary value of g are

first (see Bliss [6], p 203)

$$(15) \qquad \text{rank} \begin{Bmatrix} \partial h_n/\partial x & \partial h_n/\partial y & \partial h_n/\partial t \\ \partial g/\partial x & \partial g/\partial y & \partial g/\partial t \\ \lambda & \mu & -\lambda\dot{x} - \mu\dot{y} \end{Bmatrix}_{t=T} \begin{matrix} \} (N\times 3) \\ = N+1 \\ \\ \end{matrix}$$

and second, the rank of the matrix obtained by omitting either of the last two rows must also be $N+1$.

The first N rows of the above matrix are the coefficients of $\Delta x, \Delta y$, δT in Δh_n, the next row are from Δg, and the last row is the corresponding set of coefficients in the differential formula (6) rewritten in the form

$$(16) \qquad [\lambda\Delta x + \mu\Delta y - (\lambda\dot{x} + \mu\dot{y})\delta T]_{t=T} = \int_0^T \vec{\Lambda}\cdot\vec{v}_p \delta p \, dt.$$

In the problem studied first, this reduces to the condition $(\lambda\dot{x} + \mu\dot{y})_T \neq 0$, and is always satisfied.

One way to determine the path is to guess a, T as before. The correctional routines may be obtained from

$$\Delta h_n = \partial h_n/\partial x \Delta x + \partial h_n/\partial y \Delta y + \partial h_n/\partial t \delta T,$$

and it may be necessary to use

$$\begin{cases} \Delta\lambda(T) = \dot{\lambda}(T)\delta T + (\partial\lambda/\partial a)_T \delta a \\ \Delta\mu(T) = \dot{\mu}(T)\delta T + (\partial\mu/\partial a)_T \delta a. \end{cases}$$

The differentials are chosen to drive residual errors to zero.

Example. Consider the problem of getting to any point where x assumes a specified value X, in minimum time, with no constraint on y. The matrix of (15) is then

$$\begin{Bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ \lambda & \mu & -(\lambda\dot{x} + \mu\dot{y}) \end{Bmatrix}.$$

Hence $\mu(T) = 0$. This yields one relation

$$\Delta\mu = (-\mu_1 \sin a + \mu_2 \cos a)\delta a + \dot{\mu}(T)\delta T = -\mu(T)$$

for $\delta a, \delta T$; the equation for Δx in (13) is the other equation needed.

An alternate procedure is to continue the computation until $x = X$; this determines T_i. Then a_{i+1} is obtained by setting $\mu(T) = \mu_{1,i} \cos a_{i+1} + \mu_{2,i} \sin a_{i+1} = 0$. It seems that the only ingenuity required is in guessing starting values and in setting up correction routines from (15), typical computing problems. As this example shows, the correction routines are not unique and some may converge better than others.

6. Comments.

In a simple example requiring about 100 time steps, each iteration took about one second and a path was obtained in from four to twenty seconds on the CDC 1604, depending on the function v and the initial guesses; this could well be decreased if desired. On the other hand, the use of empirical data and the attendant calculations will undoubted increase the time.

There are two other possible methods of solution of this problem. If the function v is independent of t, then the order of the system may be reduced by eliminating the time t. The resulting Euler equation is of order two and may be solved by a relaxation procedure (Haltiner, Hamilton, Arnason [3]). A limited comparison suggests that the differential correction procedure converges more rapidly if the number of time steps is large. No way is seen to extend the relaxation method to differential systems of higher order. The only other

method goes by various names: steepest descent or ascent, and gradient. It is based on a corrective routine similar to the constructive proof of the fundamental lemma of the calculus of variations (Courant [1] p 200). Discussions suggest that all have uncertainties associated with convergence.

Corner condition. If there are any corners (points of the curve where $\vec{v}$ is discontinuous) then the differential formulas (6) and (15) need a term added $-[v(\lambda \cos p + \mu \sin p)]_{t_1-}^{t_1+} \delta t_1$ for each point. For the values of λ, μ associated with an extremal, this vanishes (but does not vanish generally if $\lambda = \lambda_1$, $\mu = \mu_1$, etc.). A program is needed to check for this condition, if it may occur, since the values of $p(t_1-)$, $p(t_1+)$ are not close to each other. Implicit in the derivation of the above correction is that the transients introduced by "coming about" are negligible.

Many important conditions which must be checked to ensure that the resulting curve affords a minimum must be neglected in a short paper. These are covered completely, though tersely, by Bliss ([6] chapters 7,8). Many of these have no significance until a path is obtained.

As applied to <u>to ship routing</u>, these have also been discussed by de Jong [4] whose dissertation contains a discussion of several interesting cases of particular interest in air navigation. The direction perpendicular to $\vec{\Lambda}$ defines the curves of constant time associated with the extremals from the origin. An alternate method of determining minimum courses is to determine these **isochrones**, as done by Hanssen and James [5], except to use extremals and the transversal conditions. This may be done

on the computer, as outlined in some preliminary reports [10] and sections 8 and 10 of this paper.

This method of solution is an application of the methods which Bliss [8] introduced for calculating differentials in ballistics (see also Bliss [9], Ch. V for summary, and p. 125 for other references).

References

[1] Courant, R. and D. Hilbert, Methods of Mathematical Physics, New York, 1953.

[2] Pontriagin, L. S. "Optimal control processes", Automation Express, 1, 15-18, 26-30, 1959.

[3] Haltiner, G. J., H. D. Hamilton and G. 'Arnason, "Minimal-time ship routing", J. Applied Meteorology, 1, 1-7, March 1962.

[4] de Jong, H. M.,"single-heading and shortest-time navigation", Staatsdrukkeriu, 's-Gravenhage, 1953.

[5] Hanssen, G. L. and R. W. James, "Optimum ship routing", J. Inst. of Navigation, XIII, No. 3, 253-272, July 1960.

[6] Bliss, G. A., Lectures on the Calculus of Variations, Chicago, 1946.

[7] Faulkner, F. D., "Optimum ship routing II", U. S. Naval Postgraduate S hool Research Paper No. 33, June, 1962.

[8] Bliss, G. A.,"The use of adjoint systems in the problem of differential corrections for trajectories", J. U. S. Artillery 51, 296-311, 1919.

[9] Bliss, G. A., Mathematics for Exterior Ballistics, New York, 1944.

[10] Faulkner, F. D., 1."Preliminary report on ship routing theory", 16 June 1961. 2."Preliminary report on minimum-time ship routes when the speed varies with time", 21 August 1961. Available from author.

Frank D. Faulkner

Frank D. Faulkner 5 June 1962
Monterey, Calif.

Part II OPTIMUM SHIP ROUTING

Introduction

Various other aspects of optimum ship routing are taken up here. The next sections treat the case where the velocity field does not change significantly with time, and may be considered constant over short periods. The differential equations are then of lower order and the problem is simpler. Two methods of solution are discussed. Each makes use of extremals, which are the minimum-time curves, or brachistrochrones, at least over short distances. These may be used together with a relation known as "transversality" to determine equal-time curves, or isochrones. These define the boundary to the points which the ship can reach at any particular time: it can generally reach any point up to and on these equal-time curves, but no point beyond them at the corresponding time; also points on the isochrones can be attained only by extremals. The shortest-time route may be determined from these. A second method of solution is given for finding the shortest-time route directly without using the isochrones.

Next the problem of determining the equal-time curves is taken up for the case where the velocity field varies rapidly or significantly with time. Then some routines are given for effecting rendezvous between two ships, (a) when one ship is following a known course, and (b) when the two ships cooperate. Finally a brief discussion is given of problems which may be encountered in the computation.

Constant Velocity Field, Euler Equations, Transversals.

In this section it will be assumed that the speed v is a function of position and heading, but not time, $v = v(x,y,y')$, where x,y are position coordinates, and $y' = dy/dx$. For simplicity, we will generally choose the coordinate set so that the initial point is the origin and the final point $(X,0)$ is on the x axis.

The Euler equation, which is a necessary condition for minimum time will be derived now. If x,y are cartesian coordinates, the time required to go a short distance along a curve is approximately

(17) $$\Delta t = \Delta s/v = \sqrt{1+y'^2}\,\Delta x/v ,$$

and the time to go along any selected curve is then

(18) $$T = \int_0^X f(x,y,y')dx;$$

this defines $f = \sqrt{1+y'^2}/v$. If x,y are not cartesian coordinates, there are corresponding relations, depending on the metric of the coordinate set.

Let us consider two neighboring paths. On the second,

(19) $$\begin{aligned} y &\rightarrow y + \delta y \\ y' &\rightarrow y' + \delta y', \end{aligned}$$

where $(\delta y)' = \delta(y')$, and "$\rightarrow$" means "is replaced by". The integral $\int_0^X |\delta y'|\,dx$ is "small". The difference in the time required to follow the two courses may be approximated then

(20) $$\delta T = \int_0^X (f_y \delta y + f_{y'} \delta y')dx + f(x,y,y')_{x=X}\delta X ;$$

subscripts in the integrand indicate corresponding partial

derivatives. Let us integrate this by parts to eliminate $\delta y'$ from the integrand

(21) $$\delta T = \int_0^X (f_y - df_{y'}/dx)\delta y \, dx + f(x,y,y')_X \delta X + [f_{y'}\delta y]_0^X .$$

If the end points are fixed, then $\delta X = \delta y(0) = \delta y(X) = 0$. If T is a minimum, because of the choice of path, then δT must vanish for all allowed functions δy. Since δy is arbitrary except for being "small" and vanishing at the endpoints of the curve, it follows that

(22) $$\frac{d}{dx} f_{y'} - f_y = 0.$$

This is the well-known Euler equation and is a necessary condition (assuming that f has continuous second derivatives and $f_{y'y'} \neq 0$); see Bliss [6], p. 11f for a proof. We may rewrite this as

(23) $$y'' = \frac{1}{f_{y'y'}}(f_y - y'f_{y'y} - f_{y'x}),$$

which is of the form $y'' = F(x,y,y')$. It may also be rewritten as a pair of first-order equations, by introducing a new variable

(24) $$\begin{aligned} y' &= z \\ z' &= F(x,y,z), \end{aligned}$$

which is convenient for computations. It should be observed that since (23) is of second order, there are two constants of integration and an extremal is determined if the values of y, y' are given for some value of x; since $y(0) = 0$ in our problem, there is a single constant of integration to be determined.

The curves defined by (22) or (23) are called extremals.

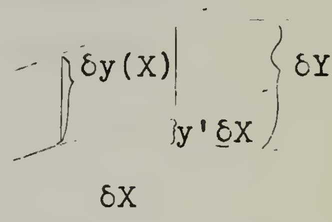
Transversals. The equations for curves which correspond to equal times along different extremals of a family, such as

the family passing through a given point, will be derived now.

If the endpoint (X,Y) is not fixed, but may vary, then the terms in equation (21) associated with it do not necessarily vanish. Since we are interested in shortest-time routes, we will consider only extremals, so that the integrand and the integral vanish. Now the change in the final value of y is

(25) $\delta Y = \delta y(X) + y'(X)\delta X$

(see sketch), y' being the value on the extremal. Equation (21) becomes



Relations among variations of end values.

(26) $\delta T = [f\delta X + (\delta Y - y'\delta X)f_{y'}]_X$

Now if we have a one-parameter family of extremals such as those coming from a given point, then the various points, one on each extremal, which correspond to the same value of T will be obtained by setting $\delta T = 0$. That is, from (25) and (26),

(27) $\delta Y - [y' - f/f_{y'}]_X \delta X = 0;$

this is the defining relation for the <u>transversal</u> direction $\delta Y/\delta X$. A curve S which cuts each of a family $\{E\}$ of extremals transversally is transversal to the family, or is a transversal of the family. Its equation is

(28) $dY/dX = y' - f/f_{y'}$;

dY/dX is the slope of S and y' is the value on the corresponding extremal.

And if we are given a smooth curve S, then there is a family of extremals $\{E\}$ which satisfy (28). The family

so defined represents the curves of shortest time from the transversal S, at least if the distances are not too great.

Comment. Sometimes the transversals are perpendicular to the extremals. It may be easily verified that this is the case for an arbitrary family of extremals if and only if f has the form $f = g(x,y)\sqrt{1+y'^2}$; in the ship-routing problem this is the case if v is independent of the heading.

8. Numerical routine.

A method of determining a minimum-time ship course is now given. It is similar to the method described by Hanssen and James [5] (pp259,260) except that it makes use of the results just derived, involving extremals and transversals.

Let us take the set of extremals which emanate from the starting point. Calculate a family of these with the initial heading angle as a parameter, say taking values of $\tan^{-1}y'(0)$ one degree or ten degrees apart, using the Euler equation. Continue these for say six or twenty-four hours. The value of t must be obtained by integration and the terminal point will involve interpolation. At the endpoints determine also, from (28) the transveral direction.

We now have a set of points on the equal-time curve and the corresponding direction. Fit a curve, using this data. It is not the most common type of data since not only is the point given, but also the slope. Name this curve S_6, or S_{24}, as the case may be. It represents the maximum distance

the ship can be at that time, based on the initial velocity field.

Now revise or update the velocity field to correspond to the values forecast for six or twenty-fours ahead. Then take the one-parameter family of extremals which cut S_6 , or S_{24}, transversally. The slopes of each extremal is given by equation (28), with revised values for v, for the continuing extremals. As many extremals as desired can be drawn out from S_6. With each extremal that is continued from S_6 associate the corresponding value of $y'(0)$.

Continue these out for another six or twenty-four hours and generate another transversal S_{12}, or S_{48}. Update the field and continue until a transversal hits the desired terminal point. Interpolate to get the initial heading and the route.

This method has the desirable feature that the results may be easily interpreted, particularly by those already calculating minimum time routes by present methods. On the other hand it requires unnecessary computing to determing one route.

9. Alternate numerical routine.

It is felt that the following method will generally determine a route more quickly.

Let us consider equations (24). The equations for the variations of an extremal are

$$\begin{aligned} \delta y' &= \delta z \\ \delta z' &= F_y \delta y + F_z \delta z. \end{aligned} \tag{29}$$

Let us guess an initial value for y'. Then compute the

corresponding extremal E^0, which will probably not go through the desired endpoint $(X,0)$. Simultaneously, compute a solution to (29), with $\delta y(0) = 0$ and any convenient value of $\delta z(0)$, since the system is homogeneous and linear in $\delta y, \delta z$. The value of $y(X)$ will be in error, say it is Y^0. Treat $\delta y(X)$ as a difference; set $\delta y(X) = -Y^0$ and solve for $\delta z = \delta y'(0)$. This gives a corrected value for $y'(0)$. Continue this until the terminal value differs from the desired value of y by less than some preassigned number, the convergence criterion.

The values for the velocity field could be updated regularly every six hours as before, if desired. A limited number of computations suggests that this will give a solution quickly.

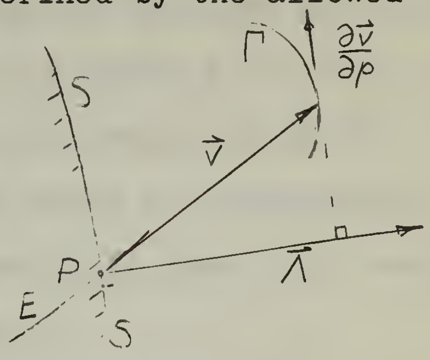
Comment. There is a condition for an extremal to yeld a minimum value for the integral, called the non-tangency condition. If at any point δy as obtained from (29) above is zero, the corresponding point is on the envelope of the family of extremals and does not furnish a minimum value to the integral. In this case a warning should be given to the operator, or a corresponding subroutine generated. This will probably never occur for short courses.

10. Isochrones for time-varying field.

A method is given here for constructing the isochrones when the velocity field varies with time, and for determining the minimum time course by the method of section 8.

Let us consider various extremals starting from the origin; each is determined by a value of a in equation (10), and by equations (1),(5),(8). At every point of each extremal E

the relation between the velocity $\vec{v}$, the solution $\vec{\Lambda}$ to the adjoint, the extremal E and the isochrone S is as indicated in the figure. Let Γ be the curve defined by the allowed velocity $\vec{v}$ at P, with p as a parameter. Then the heading p is chosen so that $\vec{v}$ has a maximum projection onto $\vec{\Lambda}$, see section 2. The curve S perpendicular to $\vec{\Lambda}$ at P is generated by small changes in the parameter a; it is a transversal of the family. The important property of S is that all points on it and to its left in some neighborhood can be reached at that time, no points to the left can be, and points of S are attained only by extremals. That is, S is part of the boundary of a closed region whose points are exactly the points where the ship may be at that time t.



Interrelations between extremal E, velocity $\vec{v}$, solution $\vec{\Lambda}$ to adjoint, and the isochrone S.

We then get points on a curve S and the tangent direction. These define the isochrone S(t). A relatively small number of points is required to define S since the tangent direction is also given; this tends to be offset by the fact that v is given from empirical data. If t is small enough, the curve S is similar to a circle or an arc thereof. If the isochrone S is given, it must be emphasized that the normal to it determines $\vec{\Lambda}$, and that the extremals are not generally perpendicular to S.

The first isochrone S(t) which touches the specified terminal

point obviously yields the minimum-time route. The value of T and the parameter a must be determined by interpolation. If the total time exceeds the time that forecasting can be done reliably, then the extremals may be replaced by terminal segments of great circle routes, the extremals for a uniform velocity field, as done by Hanssen and James [5] (p 261), or statistical data may be used as the basis for determining v. Whatever data is "most reliable" should be used; the results can be no more reliable than the data.

11. Rendezvous between ships.

Two ships will be said to rendezvous if at some time their positions coincide. It is assumed that the time required for the terminal maneuvering is negligible compared to the total time spent.

In the first example, let the position of the second ship be denoted by x^*, y^*, these being known functions of time. As before (see section 4), we guess values for the parameters of the extremal, a, T. In the correction routine for these we must allow for the distance the second ship will travel if we change T by an amount δT. In place of equations (13),(14), use

$$\begin{aligned} x^* - x &= (\dot{x} - \dot{x}^*)\delta T - \mu J \delta a \\ y^* - y &= (\dot{y} - \dot{y}^*)\delta T + \lambda J \delta a, \end{aligned} \tag{29}$$

for corrections $\delta a, \delta T$ to the estimates of a, T, all quantities in (29) being evaluated at $t = T$ after the first estimate of the trajectory has been calculated. J was defined by equation (12).

The corrected values for a, T are used and another course computed. This is continued until the distance between the two is less than some preassigned number, at the end of a computation.

Cooperating ships. The problem is more interesting if the second ship cooperates. Let quantities associated with the second ship be denoted by an asterisk (*). For rendezvous, there must be a time T such that $x(T) = x^*(T)$, $y(T) = y^*(T)$. There is a further condition that

(30) $\vec{\Lambda}(T) \parallel \vec{\Lambda}^*(T)$;

the vectors defined by the adjoint systems must be parallel at time T. That is, if $H = (\lambda\mu^* - \lambda^*\mu)_T$ then

(31) $\quad H = 0.$

We will need also the differential change in λ, μ in our computational routine. If we change a, T by small amounts, we get a differential change in the terminal values

(32)
$$\begin{aligned} \Delta\lambda &= [-\lambda_1(T)\sin a + \lambda_2(T)\cos a]\delta a + \dot{\lambda}(T)\delta T \\ \Delta\mu &= [-\mu_1(T)\sin a + \mu_2(T)\cos a]\delta a + \dot{\mu}(T)\delta T \end{aligned}$$

and two more, exactly like these, with starred terms, associated with the second ship.

Computational routine. Guess a, a^*, T and calculate a first approximation to the courses for the two ships. Let the values at the end of the i'th iteration be x_i, y_i, x_i^*, y_i^*. Generally $x_i \neq x_i^*$, or $y_i \neq y_i^*$, or $\vec{\Lambda}$ is not parallel to $\vec{\Lambda}^*$ at $t = T$. Use the differential formulas as differences to

drive the errors or residuals to zero:

$$
(33)\quad
\begin{aligned}
x_1^* - x_1 &= (\dot{x}_1 - \dot{x}_1^*)\delta T - J\mu\delta a + J^*\mu^*\delta a^* \\
y_1^* - y_1 &= (\dot{y}_1 - \dot{y}_1^*)\delta T + J\lambda\delta a - J^*\lambda^*\delta a^* \\
-(\lambda\mu^* - \mu\lambda^*)_1 &= (\dot{\lambda}\mu^* + \lambda\dot{\mu}^* - \dot{\mu}\lambda^* - \mu\dot{\lambda}^*)\delta T \\
&+ [(-\lambda_1 \sin a + \lambda_2 \cos a)\mu^* - (-\mu_1 \sin a + \mu_2 \cos a)\lambda^*]\delta a \\
&-[(-\lambda_1^* \sin a^* + \lambda_2^* \cos a^*)\mu - (-\mu_1^* \sin a^* + \mu_2^* \cos a^*)\lambda]\delta a^* .
\end{aligned}
$$

If this converges, it yields a stationary value of T. The computation would be continued until some convergence criterion is satisfied, say until $(x_1-x_1^*)^2 + (y_1-y_1^*)^2 + H^2 < \varepsilon$, where ε is a preassigned number.

An alternate procedure is the following. Suppose the coorinates are chosen so that the first ship is initially at $(0,0)$ and the second at $(X,0)$, with $X > 0$. Guess a, a^* as before and compute the two extremals, stopping when at some time $t = T_0$, $x = x^*$. Equations (33) again constitute a set of three equations for three unknowns, though only two of these, $\delta a, \delta a^*$, are needed to start the next iteration.

An alternate method would be to determine the isochrones $S(t), S^*(t)$ for each ship, by the method described in section 10, continuing until they touch. Since (or if) they are smooth curves, S and S^* are tangent to one another at the first time of contact. This is equivalent to condition (31), that $\vec{\Lambda} \parallel \vec{\Lambda}^*$, since $\vec{\Lambda}$ and S are always perpendicular.

12. Comments

As remarked initially the first purpose of this paper has been to give a method for determining optimum ship routes. It may have given however an impression that the problem is simpler than it actually is. In this section a brief discussion is given of some situations which may arise, which are not treated earlier.

It should be remarked that the routine given converges to a stationary value of time, which may not be a minimum. This is particularly the case for long routes, where a course not near that chosen may yield a smaller value of time. There is some condition which is an extension of the envelope condition in elementary calculus of variations, and the well-known conditions of Weierstrass and Legendre (see Bliss [6], chapter 8). This does not lessen the value of this paper, since these conditions can be checked only after a course has been determined.

The envelope condition in the simplest case states that if an extremal furnishes an extreme value to an integral, then it cannot contain a focal point or a point of the envelope of the family of extremals from the initial point. For the simplest problems this can be checked by checking to see if δy is zero at any point of the course, but no simple check is known generally to the author.

If the course is determined by the method of isochrones, then it is suspected that the isochrone will then develope an interior corner. Unfortunately the method of fitting the isochrones will probably have an implicit assumption of smoothness which may well obscure the existence of the corner. The difficulty is aggravated

by the fact that the data is given empirically and we must try to smooth out or minimize the random errors by a curve-fitting or smoothing routine.

For the initial part of the course we will surely be able to predict accurately. For the later part, we may be required to use statistical data, or in the absence of that, a terminal section consisting of part of a great circle. If the times involved are large, so that better data becomes available, the course should be continually redetermined, using the current position as the starting point. This is particularly important on long routes, and when storms introduce large changes in the speed which cannot be estimated accurately beforehand.

It seems that time may well be too simple a cost function. Cargo damage, danger, passenger discomfort, etc., increase with wave height. The cost function must be made up by someone familiar with the details of shipping. In the best routing procedures, several activities must be coordinated. (1) First, the meteorologist or oceanographer must collect the data and predict sea state, and weather, too, if it is significant. (2) A cost engineer familiar with the ship and its cargo must make up a realistic cost function in terms of time, danger, etc. (3) The mathematical framework is given here. (4) Finally, there is no little problem of programming properly. Some study needs to be made of methods of smoothing the data, so that the derivatives, which are notoriously poor when the data is poor, are satisfactorily smooth. The success of the method depends on the integrated efforts of all of these.

Finally, let us consider a problem where the above routine breaks down. Let us consider a sailboat in a narrow channel where hills on either side slow down the wind. Let us consider speed in the form $v = v_o(1 - y^2/w^2)(1 - 2 \cos p)$, where v_o and w are constants. It is easily verified that the maximum speed upwind (in the direction of the positive x axis is $v_o/4$. However, to effect this, the boat must come about an infinite number of times; this is known in control theory as the"chattering" solution. The above-given routine is too simple for this problem: the time spent coming about and the ensuing transients are not negligible. It seems to the author that some feel for the mechanics of such problems is necessary in the programmer, that one must suspect beforehand that such events are likely. These seem to be of little importance with powered ships and were not investigated in any detail.

Acknowledgement. This work was sponsored by the Fleet Numerical Weather Facility (FLENUMWEAFAC), Monterey, Calif., through Stanford Research Institue. I wish to express my thanks to David McIlhenny of FLENUMWEAFAC for many helpful discussions and to Mary Haynes of the Postgraduate School Computing Center for programming and assistance in checking various routines.

Frank D. Faulkner
Frank D. Faulkner
Monterey, Calif.
15 September 1962

Appendix. In this section the differential formulas of section 3 are derived. Let $\vec{\Lambda}_1, \vec{\Lambda}_2$ denote as before the solutions to the adjoint system which have initial values $\vec{i}, \vec{j}$, resp. Let $\vec{\Lambda} = \vec{\Lambda}_1 \cos a + \vec{\Lambda}_2 \sin a$, as before and $\Lambda = |\vec{\Lambda}|$. Then $\Lambda^2 = \Lambda_1{}^2\cos^2 a + \Lambda_2{}^2 \sin^2 a + 2\, \vec{\Lambda}_1 \circ \vec{\Lambda}_2 \sin a \cos a$. Let $\dot{x} = V\cos p$, $\dot{y} = V \sin p$, and denote the "velocity" by $\vec{V}$; lower case $\vec{v}$ was used earlier. Lower case symbols will be used to denote unit vectors in this section, $\vec{V} = V\vec{v}$, $\vec{\Lambda} = \Lambda\vec{w}$. Then

$$\vec{V}_p = V_p\vec{v} + V\vec{n},$$

where $\vec{n} = \vec{k}\times\vec{v}$. If $\vec{u}$ is the unit vector parallel to $\vec{V}_p$, then

$$\vec{u} = \vec{k} \times \vec{\Lambda}/\Lambda$$

and

$$\vec{V}_p = \sqrt{(V_p{}^2+V^2)}\ \vec{u}$$

on every extremal, since $\vec{V}_p \circ \vec{\Lambda} \equiv 0$.

Now let us consider at the same time and place two extremals differing by small amounts $\delta p, \delta a$. Then

$$\text{(A.1)} \qquad \vec{V}_{pp}\cdot\vec{\Lambda}\ \delta p + \vec{V}_p\cdot\vec{\Lambda}_a \delta a = 0.$$

Now also

$$\vec{V}_{pp} = |\vec{V}_p|_p\vec{u} + |\vec{V}_p|\vec{k}\times\vec{u}$$

$$= |\vec{V}_p|_p\vec{k}\times\vec{w} - |\vec{V}_p|\vec{w} \qquad\qquad {}^1\|\ \circ\Lambda\vec{w}$$

So

$$\text{(A.2)} \qquad \vec{V}_{pp}\circ\vec{\Lambda} = -\sqrt{(V^2+V_p{}^2)}\ \Lambda.$$

Now consider also

---

[1]The terms to the right of the parallel bars $\|$ in the right-hand margin will be used to suggest the operations by which the next equation is obtained.

$$\vec{V}_p\cdot\vec{\Lambda} = 0 = \vec{V}_p\circ(\vec{\Lambda}_1\cos a + \vec{\Lambda}_2\sin a)$$

Let also

$$\rho^2 = (\vec{V}_p\circ\vec{\Lambda}_1)^2 + (\vec{V}_p\circ\vec{\Lambda}_2)^2 \;;$$

then

$$\sin a = -\vec{V}_p\circ\vec{\Lambda}_1/\rho \;, \quad \cos a = \vec{V}_p\circ\vec{\Lambda}_2/\rho$$

and

(A.3) $$\vec{V}_p\cdot\vec{\Lambda}_a = \vec{V}_p\circ(-\vec{\Lambda}_1\sin a + \vec{\Lambda}_2\cos a)$$
$$= \rho.$$

Hence from (A.1),(A.2),(A.3)

$$\delta p = \rho\delta a/(\Lambda\sqrt{[V^2+V_p{}^2]}),$$

and the differential formula (6) derived earlier becomes when $\vec{\Lambda} = \vec{\Lambda}_1$

$$(\lambda_1\delta x + \mu_1\delta y)_T = -\int_0^T(\rho^2\sin a/\Lambda)dt\,da$$

for extremals. But

$$\rho^2 = [\sqrt{(V^2+V_p{}^2)}\frac{\vec{k}\times\vec{\Lambda}}{\Lambda}\circ\vec{\Lambda}_1]^2 + [\sqrt{(V^2+V_p{}^2)}\frac{\vec{k}\times\vec{\Lambda}}{\Lambda}\circ\vec{\Lambda}_2]^2$$

$$= \frac{V^2+V_p{}^2}{\Lambda^2}\Big\{[(\vec{\Lambda}_1\cos a + \vec{\Lambda}_2\sin a)\circ\vec{\Lambda}_1\times\vec{k}]^2$$
$$+ [(\vec{\Lambda}_1\cos a + \vec{\Lambda}_2\sin a)\circ\vec{\Lambda}_2\times\vec{k}]^2\Big\}$$

$$= \frac{V^2+V_p{}^2}{\Lambda^2}\begin{vmatrix}\lambda_1 & \lambda_2\\ \mu_1 & \mu_2\end{vmatrix}^2.$$

Hence

$$(\lambda_1\delta x + \mu_1\delta y)_T = -\int_0^T\frac{V^2+V_p{}^2}{\Lambda^3}\begin{vmatrix}\lambda_1 & \lambda_2\\ \mu_1 & \mu_2\end{vmatrix}^2 dt\,\sin a\,\delta a\;,$$

and

$\| -\mu_2$

$$(\lambda_2\delta x + \mu_2\delta y)_T = \int_0^T\frac{V^2+V_p{}^2}{\Lambda^3}\begin{vmatrix}\lambda_1 & \lambda_2\\ \mu_1 & \mu_2\end{vmatrix}^2 dt\,\cos a\,\delta a\;.$$

$\| \mu_1$

Hence

$$\delta x(T) = - J\,\mu(T)\,\delta a$$

$$\delta y(T) = J\,\lambda(T)\,\delta a,$$

the values of λ, μ being those associated with the maximizing condition and

$$J = \frac{1}{\begin{vmatrix} \lambda_1 & \lambda_2 \\ \mu_1 & \mu_2 \end{vmatrix}_T} \int_0^T \frac{V^2 + V_p^{\,2}}{\Lambda^3} \begin{vmatrix} \lambda_1 & \lambda_2 \\ \mu_1 & \mu_2 \end{vmatrix}^2 dt \;.$$

This derivation seems long and complex but no simplification has been found, and the corresponding three-dimensional formula has so far defied derivation.

Frank D Faulkner

Frank D. Faulkner   5 June 1962
Monterey, Calif.